

MICHIGAN STATE UNIVERSITY
Department of Statistics and Probability

COLLOQUIUM

Yuting Wei
Carnegie Mellon University

Reliable hypothesis testing paradigms in high dimensions

Tuesday, October 14, 2019
10:20 AM - 11:10 AM [Eastern Daylight Time \(EDT\)](#)
Zoom

Abstract:

Modern science and engineering applications that involve massive data urge for the design of trustworthy inferential procedures. In this talk, I will discuss two recent stories on conducting reliable inference for high-dimensional data.

The first one considers performing inference based on the Lasso estimator in the regime where the number of covariates is of the same order or larger than the number of observations. Classical asymptotic normality theory does not apply to this model due to two fundamental reasons: (1) The regularized risk is non-smooth; (2) The discrepancy between the estimator and the true parameter vector cannot be neglected. As a consequence, standard perturbative arguments that are the traditional basis for asymptotic normality fail. In this work, we study the distribution of the Lasso as well as its debiased version and show that a degrees-of-freedom correction is necessary for computing valid confidence intervals.

In the second part, we investigate the Model-X knockoffs framework, which is a general procedure that can leverage any feature importance measure to produce a variable selection algorithm. Since it relies on the construction of synthetic random variables, it is, therefore, random. In this work, we propose a method for derandomizing model-X knockoffs. By aggregating the selection results across multiple runs of the knockoffs algorithm, our method provides stable decisions without compromising statistical power. Our approach, when applied to the multi-stage GWAS of prostate cancer, reports locations on the genome that have been replicated with other studies.

This talk is based on joint work with Michael Celentano, Andrea Montanari, Zhimei Ren, and Emmanuel Candes.

Bio:

Yuting Wei is currently an assistant professor in the Statistics and Data Science department at Carnegie Mellon University. Prior to that, she was a Stein Fellow at Stanford University and she received her Ph.D. in statistics at University of California, Berkeley working with Martin Wainwright and Aditya Guntuboyina. She was the recipient of the 2018 Erich L. Lehmann Citation from the Berkeley statistics department for her Ph.D. dissertation in theoretical statistics. Her research interests include high-dimensional and non-parametric statistics, statistical machine learning, and reinforcement learning.

Zoom details can be found at: <https://stt.natsci.msu.edu/stt-colloquium-zoom-info/>

To request an interpreter or other accommodations for people with disabilities, please call the Department of Statistics and Probability at 517-355-9589.